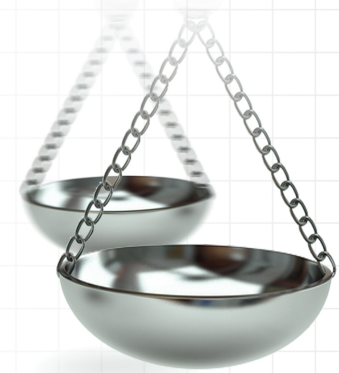
Methodology

For reprint orders, please contact: reprints@futuremedicine.com



Journal of **Comparative Effectiveness Research**

Developing a cross-validation tool for evaluating economic evidence in rapid literature reviews

Xuanqian Xie*[,1], Chunmei Li[1], Sean Tiggelaar[1], Frances Simbulan[1], Lindsey Falk[1] & James M Brophy[2]

[1]Health Technology Assessment Program, Ontario Health, Toronto, ON M5S 1N5, Canada

[2]Departments of Medicine & Epidemiology, Biostatistics, & Occupational Health, McGill University, Montreal, QC H3A 0G4, Canada

*Author for correspondence: Tel.: +1 647 953 1322; shawn.xie@ontariohealth.ca

Background: Rapid economic reviews efficiently summarize economic evidence. However, reporting main findings without assessing quality and credibility can be misleading. The objective of this study was to develop a rapid cross-validation screening tool to evaluate economic evidence when conducting rapid economic literature reviews. **Methods:** This article outlines our reasoning and the theoretical concepts for developing the screening tool. **Results:** This cross-validation tool is a qualitative approach under a Bayesian framework that uses prior health economic evidence to gauge the credibility of the rapid economic review's findings. This article describes an application of this tool and highlights practical considerations for its development and deployment. **Conclusion:** This tool can provide a valuable screening instrument to evaluate the quality and credibility of the economic evidence.

First draft submitted: 2 November 2021; Accepted for publication: 9 September 2022; Published online: 28 September 2022

Keywords: Bayesian framework • cross-validation tool • economic evidence • QALY • rapid review

Background

When deciding whether to adopt a new health technology, it is critical to identify and evaluate systematically the published economic evidence to make informed decisions regarding value. However, it is not always feasible to conduct a full health technology assessment (HTA) and a resource-intensive systematic review. In recent years, rapid reviews (i.e., rapid HTA, mini HTA and rapid response) have been extensively used throughout many HTA agencies, including the National Institute for Health and Care Excellence (NICE) in the UK and the Canadian Agency for Drugs and Technologies in Health (CADTH) in Canada [1,2]. Specifically, CADTH adopted its rapid review program in 2005 and has since published hundreds of rapid reviews annually [1].

This shift in focus has led to different challenges. It is challenging to synthesize and assess the quality of economic evidence in a rapid HTA due to constrained timelines, increased complexity of economic models and lack of transparency because most model-based economic evaluations do not publish their models. Alternatively, a rapid economic literature review may be misleading without a critical evaluation of the validity of the findings, which may lack robustness or introduce bias. A systematic review on published cost–effectiveness studies, found that most published studies reported favorable cost–effectiveness results [3], a result that may be tied to ubiquitous industry sponsorship [4,5].

This article aims to outline the theory underlying the development of a screening tool that aids in critical evaluation of studies when conducting rapid economic reviews. Researchers from various settings (e.g., academia, HTA agencies and industry) who conduct rapid economic reviews may benefit from this cross-validation tool. In addition, this tool can be valuable for stakeholders who need to understand the quality and credibility of findings in rapid reviews for decision-making. Conceptually, it is a quick qualitative approach under a Bayesian framework that uses prior health economic evidence (here termed 'the reference') to gauge the credibility of the economic evidence in rapid reviews. This tool is especially useful when the standard HTA or systematic reviews are not available or feasible to conduct. The tool works by forming a small database of the reference economic evidence.

Future Medicine

Within that database it captures two main components, the clinical benefit (e.g., improving disease progression) and corresponding quality-adjusted life years (QALYs) gained. The tool focuses on the logical consistency of model input (e.g., clinical benefits) and outputs (i.e., QALYs gained for cost–utility analysis [CUA]) of findings within a rapid review, compared to the reference (i.e., prior economic evidence). This tool does not focus on the technical details (e.g., model structure, assumptions and parameters) of economic modeling. The underlying assumptions of the tool are as follows:

1) There is general consensus on the intervention's safety profile, clinical effect and the clinical effect's duration
2) A larger magnitude of clinical benefits is generally associated with greater QALYs gained
3) The best-quality economic evaluation captures the unknown 'true' health benefits in QALYs

When there is no major violation of the assumptions, appropriate analytic approaches should provide consistent results. Therefore, by identifying the degree to which the new intervention provides clinical benefits (i.e., greater, smaller or similar) compared with the existing interventions, a reasonable interpolated range of QALYs gained from the new intervention can be approximated.

The remainder of this article is outlined as follows: first, an overview of commonly used quality and validity tools and considerations for developing our new cross-validation tool is provided. The next section introduces Bayesian reasoning – the mathematical concepts of this tool – and its extensions (e.g., a simplified tool). Next, examples of using these concepts are provided. The next sections outline the practical considerations of the tool's development and deployment and discuss the findings of the tool and its limitations. Finally, conclusions are provided.

Considerations for a new tool for rapid economic literature reviews

The goals of economic literature reviews are not solely to report and summarize study findings, but also to assess their methodological quality, robustness and overall credibility to ensure that the results can be safely extrapolated for decision making. Several checklists or quality assessment tools have been developed to help evaluate the applicability (i.e., how relevant is the study to the current decision problem) and quality (i.e., risk of bias) of published economic studies [6–9]. Additionally, a structured validation assessment tool has been developed for reviewing model development [10]. In general, these tools are comprehensive with many items to examine, causing them to be cumbersome, if not inapplicable, for a rapid review. To help address these issues when conducting a rapid review, we propose an efficient tool that provides a quick assessment of studies' validity.

According to the ISPOR-SMDM Modelling Good Research Practices Task Force, model validation involves five steps that include face, internal, cross, external and predictive validity [11]. Specifically, cross-validity is when the results of one model are compared with another model answering the same research question. When using our cross-validation tool, a reference study that is deemed to be the highest quality study is selected from the list of economic studies (i.e., the robust external prior knowledge). The QALYs gained from this reference study are referred to be the standard value used to compare the QALYs gained from other studies. The reference study's results are compared against the findings from the other economic studies in the review, with the goal of validating them. Because this cross-validation tool is intended to assess economic evidence in rapid reviews, it does not assess model structure and statistical methods. This tool strictly evaluates the clinical benefits and QALYs gained for each treatment, by disease category.

Compared with health outcomes reported in natural units such as life expectancy, QALYs incorporate gains in both quantity of life (i.e., length of life) and quality of life (i.e., one QALY represents a year in perfect health). Guidelines for economic evaluations recommend the use of QALYs when possible as it allows for comparisons of different technologies across disease areas, thus supporting decision makers with evidence to allocate resources across different conditions [6,12]. As a result, QALYs is the most universal outcome measure used in health economic evaluations. This cross-validation tool does not focus on the total QALYs of an individual treatment because the absolute QALYs for a treatment can be associated with multiple factors, such as the model's time horizon and baseline risk (i.e., a model with a longer time horizon will produce larger QALYs for an intervention). Alternatively, the tool is used to evaluate the incremental difference in QALYs between two treatments because it may have a smaller variation than the total QALYs. For example, model guidelines indicate studies should use time horizons that capture the major differences in costs and outcomes associated with the interventions of interest. As a result, the QALY difference between two treatments may be very similar in a model with a 10-year time horizon or a lifetime horizon (e.g., the health outcome and survival are likely similar or practically identical after 10 years).

Given that healthcare resource use and unit costs are often region and country specific, we do not consider costs in this tool. However, we do provide some thoughts on incorporating cost and incremental cost–effectiveness ratios (ICERs) in the discussion section.

Theoretical concepts behind the cross-validation tool & its extensions

Bayesian reasoning: evaluating new findings in the context of existing evidence

This qualitative approach of evaluating economic evidence was inspired by the Bayesian framework. The key feature of the Bayesian framework is to allow new findings to be set in the proper context [13]. Briefly, Bayesian posterior distributions were derived by updating prior probabilities with new observed data. Bayesian methods naturally permit the synthesis of all available evidence (i.e., not the analysis of the new data alone); therefore, it can improve the precision of treatment effect estimates and potentially reduce the bias associated with such estimates [14].

Following Bayesian thinking, the purported QALYs gained in the rapid economic review study are put into the context of the existing health economic evidence. The best quality economic study was selected as the reference, which was assumed to reflect the true QALYs gained, to act as a benchmark for the results of other analyses (this selection process is further explained later in the article). These reference studies act as a safeguard, reducing potential biases of interpreting data from an individual economic evaluation. When the results of individual studies are inconsistent with the existing evidence, the findings of new study may be seen as lacking credibility.

Basic concepts of the cross validation tool

Using 83 published articles, Wright and Weinstein categorized the gains in life expectancy from various medical interventions as large or small. The interventions included both treatments and preventive interventions, stratified by the target population and disease [15]. For example, implantable cardioverter-defibrillators lead to a life-expectancy gain of 36 to 46 months for the survivors of cardiac arrest, whereas amiodarone therapy leads to a life-expectancy gain of 14 to 16 months. Also, there is an association between clinical outcomes and generic health related quality of life (HRQOL). Methods of mapping clinical outcomes to health utility have been developed and used in economic evaluations [16]. Thus, it is possible to evaluate, quantitatively or qualitatively, the magnitude of utility improvement by an intervention, on the basis of its impact on severity and frequencies of clinical events. We further extended these concepts to the association between the clinical benefit (set A) and the QALYs gained (set B). Researchers developing this tool must review both clinical and economic literature to populate sets A and B.

Let A be the set of clinical benefits (a_i) for treatments for a given indication. Each element (a_i) represents the magnitude of the difference in clinical effectiveness between two treatments (or between an intervention and placebo).

$$A = \{a_1, a_2, a_3, \ldots, a_n\}$$

Let us select numerous interventions whose clinical benefits can be ranked by clinical judgment or clinical outcome measures. Let $a_1 > a_2 > a_3 > \ldots > a_n > 0$. Importantly, specific numeric values do not need to be assigned for the elements in set A because this tool is qualitatively structured. To help illustrate this concept, the following is an example in heart disease. The average clinical benefits (e.g., gain in life expectancy) of heart transplantation are greater than myocardial revascularization with coronary artery bypass grafting (CABG), and the average clinical benefits of CABG are greater than radiofrequency catheter ablation (a surgery that uses radiofrequency energy to destroy a small area of heart tissue, causing rapid and irregular heartbeats) [15,17,18].

Assume that clinical benefits can be directly represented by the differences in QALYs between interventions (i.e., QALYs gained). For simplicity, further assume that clinical benefit is the main contributor to differences in QALYs between select interventions, whereas other factors (e.g., adverse events) have an insignificant impact (note: later a scenario where QALY differences between treatments is driven by the safety profile is discussed). Therefore, clinical benefits will be positively correlated with QALYs gained (i.e., a larger magnitude of clinical benefit is associated with greater QALYs gained). Let B be the set of QALYs gained (b_i).

$$B = \{b_1, b_2, b_3, \ldots, b_n\}, \text{ where } b_1 > b_2 > b_3 > \ldots > b_n > 0.$$

Each element in set B has a corresponding numeric value of QALYs gained.

Each element in set A (a_i) has a corresponding element in set B (b_i), and the order of the comparisons in set A and B should be the same or similar. Let us assume that the order of all elements in set A are known, and the values in a subset of set B are known. On the basis of the known order of elements in set A and the known values in the subset of B, we can deduce a range of median values for unknown elements in set B.

Simplified cross-validation tool

For most diseases, there are a small number of treatment options available. Therefore, it is unlikely that there will be many elements for set A and B. We can simplify this by using only two elements, one with the largest clinical benefits (e.g., a highly effective treatment that is infrequently used in clinical practice, such as heart transplants) and the other with the smallest clinical benefits (e.g., pharmaceutical therapy). Once we have two elements that represent both the largest and smallest clinical benefit, we can theoretically obtain an interval, and we expect that a new intervention's QALYs gained will likely fall within this interval. When we extend this concept to the simplest condition (only one reference treatment), then, based on the clinical judgment of the new intervention in relation to the reference, we can estimate whether the QALYs gained of the new intervention is larger or smaller than the 'true' value from the reference.

QALYs gained driven by the safety profile & HRQOL

We have discussed the relationship between an intervention's gains in QALYs being driven by the difference in clinical benefits between the interventions. For a given condition, the primary clinical effects may be similar across various interventions; however, the secondary outcomes (e.g., the safety profile) may differ. For example, different oral anticoagulants (dabigatran, rivaroxaban and apixaban) have a similar effect on the risk of ischemic stroke in patients with atrial fibrillation, but based on observational research, apixaban is likely to be associated with a lower risk of major bleeding [19,20]. In this case, the QALYs gained are driven by the safety profile. Following the concepts discussed earlier, the elements of set A can be ranked by safety, and elements in set B are the corresponding QALYs gained in this condition.

This tool may be extended to interventions that improve HRQOL but have no effect or marginal effect on survival, clinical events (e.g., stroke) or other objective measures, such as treatments for anxiety and depression [21,22]. When ranking these treatments, researchers need to consider both the HRQOL instruments used (i.e., different HRQOL instruments used for one disease) and the magnitude of HRQOL benefits from an instrument. Furthermore, although used in various diseases in practice, generic HRQOL instruments such as the EQ-5D, may be less responsive to capturing small changes in HRQOL for some conditions, such as in patients with prostate cancer [23]. For these conditions, using this tool may lead to greater uncertainty in estimating the QALYs gained for a given intervention.

When both primary and secondary outcomes have considerable contributions to the QALYs gained, the ranking of health benefits may be complex and undetermined. Therefore, this tool may be unsuitable for this scenario.

Applications for the cross-validation tool

Defining the 'true' health benefits in QALYs

The key factor in applying this cross-validation tool is to select references properly and define the true QALYs gained from an intervention. Although efficacy measures are often considered constant, the results of economic evaluations often differ because of varying contexts (e.g., varying clinical practice or different costs), model assumptions or controls of bias parameters (e.g., measurement errors, selection bias, confounding or conflict of interest). Theoretically, however, if effectiveness is estimated based on the same body of key clinical evidence, the findings from different economic evaluations for a specific intervention should be relatively consistent. Therefore, although subject to limitations, we can assume that the best quality economic evaluation for an intervention approximates the 'true' but unknown health benefit in QALYs.

A hypothetical example of using the cross-validation tool

Figure 1 shows the illustrated process of implementing the cross-validation tool. Once the tool (i.e., database of clinical benefits and corresponding QALYs gained) is established and validated, we can quickly estimate a new intervention's range of QALYs gained compared with an older treatment once the clinical effect of this intervention is established.

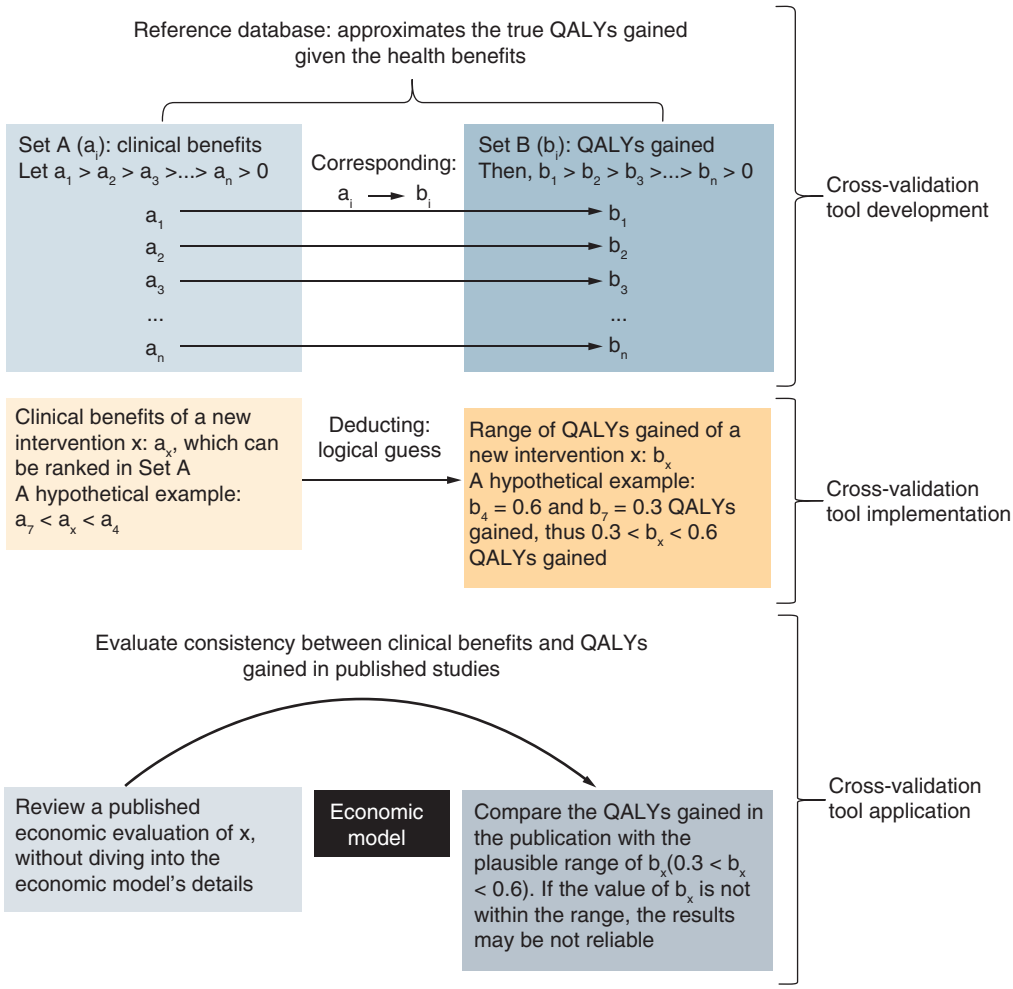


Figure 1. Implementation of the cross-validation tool.
QALYs: Quality adjusted life years.

In a hypothetical example in Figure 1, according to clinical judgement (or clinical measures) the benefits of the new intervention would be less than the fourth-ranked treatment in the dataset (associated with 0.6 QALYs gain compared with usual care) and greater than the seventh-ranked treatment (associated with 0.3 QALYs gained compared with usual care). Therefore, the QALYs gained associated with the new intervention likely range from 0.3 to 0.6. If the results of a newly published economic evaluation contradict this value, the validity of the new economic evaluation's findings may be seen as questionable.

A single reference case: CGM HTA from NICE

This idea can be further explained using a case example. Both continuous glucose monitoring (CGM) and the flash glucose-monitoring system (Flash) measure interstitial glucose and provide data on glucose trends for diabetes insulin therapy. Flash has similar functionality to CGM, and the price is significantly less. However, CGM has the added function of alarms that alert the user to hypoglycemia or hyperglycemia [24]. Thus, CGM should have either better or similar effectiveness than Flash for type 1 diabetes because CGM has more features [24]. A diabetic modeling study prepared by NICE evaluated the cost–effectiveness of CGM versus self-monitoring of blood glucose (SMBG) in patients with type 1 diabetes [25]. The results showed that compared with self-monitoring, CGM had an increase of 0.10 QALYs over a lifetime horizon (continuous subcutaneous insulin infusion [CSII] + CGM: 12.82 QALYs versus CSII + SMBG: 12.72 QALYs). If we assume the NICE study represents the true effectiveness of CGM, we can expect Flash would have a smaller or equal benefit compared with SMBG (QALYs gained of around 0.10). However, a recent CUA in patients with type 1 diabetes found that compared with SMBG, Flash had a large

gain of 2.12 QALYs over a lifetime horizon [26]. By comparing the findings with that of the NICE study, we may be curious about the validity of the economic analysis' results. When the study was appraised further, we found that it likely overestimated the QALYs gained associated with Flash because the disutility of nonsevere hypoglycemia may have been overestimated by assuming a linear relationship between the frequency of hypoglycemic events and the total reduction in utility.

Considerations in developing the cross-validation tool

This section proposes specific considerations in developing both set A (set of clinical benefits) and B (set of QALYs gained) of this cross-validation tool. With the wider application of this tool, more suitable approaches of defining set A and B would be addressed.

Considerations for developing set A

To create set A, it is critical to develop a dataset with reference treatments by disease category. Further considerations in developing set A are as follows.

Established interventions

Interventions that have well established clinical evidence to support their effect should be selected. We can select some milestone treatments (or important innovations) in the history of medicine as the references. A network meta-analysis is a possible source of identifying the interventions that have shown statistically significant benefits compared with other treatments and have relatively less uncertainty in the magnitude of effectiveness (i.e., a narrow confidence interval). If there are no direct pairwise comparisons between the interventions and comparators of interest, the health benefits of the intervention showed in the network meta-analysis would have to be evaluated by clinical judgment.

Disease categories

Although it is not impossible to compare the clinical benefits of interventions for different diseases or indications (e.g., cardiovascular diseases, oncology), it is straightforward to group the important interventions by disease categories. The disease severity level should be similar in a single dataset because the health benefits of an intervention are strongly correlated with patients' risk profile.

Types of treatments

Different types of treatments, such as surgical and pharmaceutical treatments can be in the same dataset for similar indications. However, drugs used for primary and secondary prevention (e.g., cardiovascular disease) may be in different datasets given considerable differences in baseline risk.

Diagnostic tests

In general, set A should not include diagnostic tests. This is because one test can be used in different populations for different purposes. Translating the values of diagnostic or screening tests into clinical benefits would involve considering the downstream treatments for true positive (and sometimes true negative) results, and management of false positive and false-negative results. Ultimately, we would likely exclude this because it would bring in more uncertainty to the tool.

Magnitude of benefits

The reference treatments should cover the wide range of clinical benefits (e.g., a large survival benefit from heart transplantation and a relatively small benefit from radiofrequency catheter ablation for heart disease).

Size of database

The size of the database is flexible. For certain disease conditions, there may not be many appropriate milestone treatments available to use as the references. Also, if the size of dataset is too large, it may be difficult to reach consensus when ranking all references.

Considerations of developing set B

Following are considerations for selecting the best quality economic evaluations of key treatments for a disease (i.e., selecting elements for set B).

Transparency

Although not used in the tool's evaluation, the reference studies should report details such as the model structure, model parameters, calculation process of derived parameters and assumptions. Ideally, the economic model should be published. Caution is warranted in studies with a complex model structure that do not publish their model because the results cannot be easily verified.

Conflict of interest

All coauthors should have no conflict of interest, and ideally all coauthors should be from the public sector. If there is a conflict of interest, bias may be introduced into the economic evaluations.

Quality of studies

Although most published economic evaluations have been peer reviewed, the methodological quality can vary greatly. To include published economic studies in set B as references, two or more researchers need to review these studies independently to confirm that the study is of high quality. If there are two high-quality economic studies that yield slightly different results based on the same key clinical evidence, set B may include both studies to allow variation of estimates. However, if the QALYs gained estimated by these two studies are considerably different, researchers need to further assess relevant model characteristics (e.g., comparator selection, time horizon) to understand the differences in the findings. Subsequently, researchers need to make a judgment as to which study is applicable and most relevant to the setting, scope or context of interest and keep a single unbiased estimate in set B.

Sources of main clinical parameters

The main clinical parameters of reference economic studies should be based on the best available clinical evidence, with justification for their source. For selecting the reference studies for set B, we may not include economic studies that used local data (e.g., data from administrative dataset for a particular setting) as the key clinical model inputs, given that the results from these studies may be not generalizable to other settings.

Year of publication

We should focus on economic studies published in recent years because medical evidence may update over time (e.g., newer evidence, studies with longer follow-up). Standard patient management, clinical pathways and background risk of the population (e.g., the risk of mortality for general population) may change over time, and thus studies published in recent years should provide a more reliable estimate of QALYs gained for the population of interest.

Practical considerations of developing & applying the cross-validation tool

Researchers from various settings such as academia, HTA agencies or industry that frequently conduct rapid HTAs or rapid reviews may benefit from the development of this cross-validation tool. Including health economists, researchers with medical or clinical background may contribute to this developing process to ensure the suitability of the clinical judgments for the set A. If the HTA agencies are able to conduct full-HTAs, they may use recent primary economic elevations in HTAs to develop set B, which would be relevant to the setting and context of interest. After establishing the initial version of this tool, researchers can adapt this tool to evaluate economic evidence in their rapid literature reviews and provide the feedback to continue improving this validation tool. It is expected that set A does not need to be modified often, but the estimate of QALYs gained in set B may be modified with the new estimates from recently published economic evaluations.

When applying this tool to a published economic study included in the rapid review, the researcher may provide either positive or negative comments. For example, when the QALY gained of the study is not consistent with the existing evidence, the researcher can state that 'We did not use a structured validation assessment tool to evaluate the quality of this economic study in this rapid review. However, we used a cross-validation tool to evaluate the cost–effectiveness results and found that the QALYs gained from the intervention in this study was likely beyond the possible range, based on the existing high-quality economic evidence. Therefore, the findings of this study should be interpreted with caution.'

For the individual researchers who occasionally conduct the rapid reviews, it may be impractical to develop the cross-validation tool for a single rapid review. However, they can source established tools that are available from

other HTA agencies (given that a tool is widely accepted in the future). Lastly, researchers may consider using the simplified tool, as illustrated earlier for the concepts and the example.

Limitations & discussion

Before using this tool, it is important to understand its limitations. First, ideally Sets A and B would be objective and include standardized reference interventions (e.g., standard of care, placebo). Although in practice this may be difficult to achieve, users should strive to use interventions with the most reliable and consistent evidence base. Second, although this tool assumes that such reference interventions capture the 'true' QALYs gained, this is challenging to confirm. Third, this tool does not consider the validation of costs and ICERs because they may be strongly influenced by jurisdictional variation in clinical practice, healthcare structure, costs and commonly accepted cost–effectiveness thresholds. However, this tool could potentially be extended to incremental costs and ICERs, particularly when clinical practices (e.g., guidelines), healthcare systems and costs are similar across evaluated studies. Fourth, our tool uses QALY as the measure of effectiveness, whereas other effectiveness measures are also used in economic evaluations. Fifth, this tool is based on the existing high-quality economic studies; in some disease areas, there may be lack of the existing economic evidence, and thus the tool may not be applicable in this scenario.

It is important to recognize that our tool is not intended to replace quality or validation tools currently used in systematic literature reviews [6–9] and economic model development [10,27]. However, when the true QALYs gained of a new intervention is unknown, it may be reasonable to use carefully selected treatments as references to theorize the QALYs gained of a new intervention.

Conclusion

This study presents a theoretical screening tool for rapid economic reviews that assesses the consistency of economic results using a qualitative synthesis of previously published studies. This provides researchers a valuable tool for systematic comparison of a new intervention's QALYs gained against findings from other economic evaluations that allows one to judge quickly whether the results of an economic study appear reasonable in light of the existing evidence.

Summary points

- Synthesizing and assessing the quality of economic evidence in a rapid economic literature review may pose several challenges to researchers due to constrained timelines, increased complexity of newly published economic models and lack of transparency. Alternatively, it can be misleading if a rapid review does not critically evaluate the validity of the findings, which may lack robustness or even be biased.
- This work aims to outline the theory underlying the development of a screening tool that aids with critical evaluation of economic evidence when conducting rapid literature reviews. This tool allows for an efficient qualitative approach, under a Bayesian framework, using prior health economic evidence (termed 'the reference' for the purposes of this work) to gauge the credibility of the rapid economic review's findings.
- Following Bayesian thinking, the purported QALYs gained in the rapid economic review study are put into the context of the existing health economic evidence. The best quality economic study was selected as the reference and assumed to reflect the true QALYs gained and act as a benchmark for the results of other analyses. When the results of individual studies are inconsistent with the existing evidence, the findings of a new study may be seen as lacking credibility.
- Using previously published articles, researchers can categorize the gains in life expectancy from various medical interventions as large or small. Also, methods of mapping clinical outcomes to health utility have been developed and used in economic evaluations. Thus, it is possible to evaluate the magnitude of QALYs gained by an intervention qualitatively, based on the intervention's impact on severity and frequencies of clinical events.
- The tool works by forming a small database of the reference economic evidence. Within that database, it captures two main components: the clinical benefit and corresponding QALYs gained. Therefore, by identifying the degree to which the new intervention provides clinical benefits (i.e., greater, smaller or similar) compared with existing interventions, a reasonable interpolated range of QALYs gained from the new intervention can be approximated.

Author contributions

X Xie conceived the study idea. JM Brophy supervised this study. X Xie, C Li, S Tiggelaar, F Simbulan, L Falk and JM Brophy designed the study, wrote the manuscript and interpreted the findings.

Disclaimer

The opinions expressed in this publication do not necessarily represent the opinions of Ontario Health. No endorsement is intended or should be inferred.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Khangura S, Polisena J, Clifford TJ *et al*. Rapid review: an emerging approach to evidence synthesis in health technology assessment. *Int. J. Technol. Assess. Health Care* 30(1), 20–27 (2014).
2. Kaltenthaler E, Cooper K, Pandor A *et al*. The use of rapid review methods in health technology assessments: 3 case studies. *BMC Med. Res. Methodol.* 16(1), 108 (2016).
3. Bell CM, Urbach DR, Ray JG *et al.* Bias in published cost effectiveness studies: systematic review. *BMJ* 332(7543), 699–703 (2006).

• **A systematic review showed that most published studies reported favorable cost–effectiveness results.**

4. Jang S, Chae YK, Majhail NS. Financial conflicts of interest in economic analyses in oncology. *Am. J. Clin. Oncol.* 34(5), 524–528 (2011).
5. Al-Badriyeh D, Alameri M, Al-Okka R. Cost-effectiveness research in cancer therapy: a systematic review of literature trends, methods and the influence of funding. *BMJ Open* 7(1), e012648 (2017).
6. National Institute for Health and Care Excellence. *Methods for the Development of NICE Public Health Guidance (3rd edition)*. NICE, London (2012).
7. Evers S, Goossens M, De Vet H *et al*. Criteria list for assessment of methodological quality of economic evaluations: consensus on Health Economic Criteria. *Int. J. Technol. Assess. Health Care* 21(2), 240–245 (2005).
8. Drummond MF, Sculpher MJ, Claxton K *et al*. *Methods for the economic evaluation of health care programmes*. Oxford University Press, Oxford, UK (2015).
9. Jaime Caro J, Eddy DM, Kan H *et al.* Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health* 17(2), 174–182 (2014).
10. Vemer P, Corro Ramos I, Van Voorn GA *et al*. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics* 34(4), 349–361 (2016).
11. Eddy DM, Hollingworth W, Caro JJ *et al.* Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force – 7. *Value Health* 15(6), 843–850 (2012).
12. Canadian Agency for Drugs and Technologies in Health. *Guidelines for the economic evaluation of health technologies: Canada (4th Edition)* CADTH, Ottawa, Canada (2017).
13. Matthews RaJ. Methods for assessing the credibility of clinical trial outcomes. *Drug Inf. J.* 35(4), 1469–1478 (2001).

•• **A study introduced the Bayesian framework to evaluate the credibility of clinical trial results.**

14. Labos C, Brophy JM, Sniderman A, Thanassoulis G. Mortality benefit of alirocumab: a Bayesian perspective. *J. Am. Heart Assoc.* 8(20), e013170 (2019).
15. Wright JC, Weinstein MC. Gains in life expectancy from medical interventions – standardizing data on outcomes. *N. Engl. J. Med.* 339(6), 380–386 (1998).

•• **A study showed that the gains in life expectancy by interventions can be qualitatively evaluated.**

16. Hernandez Alava M, Wailoo A, Pudney S *et al*. Mapping clinical outcomes to generic preference-based outcome measures: development and comparison of methods. *Health Technol. Assess.* 24(34), 1–68 (2020).

• **This study showed that it is possible to evaluate, quantitatively or qualitatively, the magnitude of health utility improvement by an intervention.**

17. Van Der Wall EE. Milestones in cardiovascular medicine: 10 or more? *Neth. Heart J.* 21(12), 527–529 (2013).
18. Chan PS, Vijan S, Morady F, Oral H. Cost-effectiveness of radiofrequency catheter ablation for atrial fibrillation. *J. Am. Coll. Cardiol.* 47(12), 2513–2520 (2006).
19. Douros A, Durand M, Doyle CM *et al*. Comparative effectiveness and safety of direct oral anticoagulants in patients with atrial fibrillation: a systematic review and meta-analysis of observational studies. *Drug Saf.* 42(10), 1135–1148 (2019).

20. Rutherford OW, Jonasson C, Ghanima W *et al*. Comparison of dabigatran, rivaroxaban, and apixaban for effectiveness and safety in atrial fibrillation: a nationwide cohort study. *Eur. Heart J. Cardiovasc. Pharmacother.* 6(2), 75–85 (2020).
21. Cramer H, Lauche R, Anheyer D *et al.* Yoga for anxiety: a systematic review and meta-analysis of randomized controlled trials. *Depress. Anxiety* 35(9), 830–843 (2018).
22. Letourneau NL, Dennis CL, Cosic N, Linder J. The effect of perinatal depression treatment for mothers on parenting and child development: a systematic review. *Depress. Anxiety* 34(10), 928–966 (2017).
23. Krahn M, Bremner KE, Tomlinson G *et al*. Responsiveness of disease-specific and generic utility instruments in prostate cancer patients. *Qual. Life Res.* 16(3), 509–522 (2007).
24. Garg SK, Akturk HK. Flash glucose monitoring: the future is here. *Diabetes Technol. Ther.* 19(S2), S1–S3 (2017).
25. Riemsma R, Corro Ramos I, Birnie R *et al.* Integrated sensor-augmented pump therapy systems [the MiniMed(R) Paradigm Veo system and the Vibe and G4(R) PLATINUM CGM (continuous glucose monitoring) system] for managing blood glucose levels in type 1 diabetes: a systematic review and economic evaluation. *Health Technol. Assess.* 20(17), 1–251; v–xxxi (2016).
26. Healthcare Improvement Scotland. *Evidence Note 81: What Is the Clinical and Cost Effectiveness of Freestyle Libre® Flash Glucose Monitoring for Patients With Diabetes Mellitus Treated With Intensive Insulin Therapy?* HIS, Edinburgh, UK (2018).
27. Buyukkaramikli NC, Rutten-Van Molken M, Severens JL, Al M. TECH-VER: a verification checklist to reduce errors in models and improve their credibility. *Pharmacoeconomics* 37(11), 1391–1408 (2019).